DOCUMENT RESUME

ED 431 000 TM 029 826

AUTHOR Nweke, Winifred C.; Hall, Thomas R.

TITLE Evaluating Cut-Scores on Two Certification Tests: How Well Do Decisions Based on Cut-Scores Match Teacher- and Principal-Reported Ratings of Competence in the Classroom?

PUB DATE 1999-04-00

NOTE 26p.; Paper presented at the Annual Meeting of the American Educational Research Association (Montreal, Quebec, Canada, April 19-23, 1999).

PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.

DESCRIPTORS *Beginning Teachers; *Competence; *Concurrent Validity; *Cutting Scores; Elementary Secondary Education; Principals; *Teacher Certification; *Teacher Evaluation; Teacher Qualifications

IDENTIFIERS *Georgia Teacher Certification Testing Program; Praxis Series

ABSTRACT

The purpose of this study was three-fold. The first purpose was the investigation of the criterion-related validity of the Georgia Teacher Certification Test (TCT) and the Praxis II tests that are used in the teacher certification process in Georgia. The second purpose was to compare decisions based on the two tests. Finally, the effects of using recommended, rather than the adopted, cut-scores were examined. Participants were 2,326 beginning teachers in Georgia in the 1998 fiscal year and their principals. Beginning teachers and their principals completed questionnaires that elicited, on a four-point scale, how well prepared and ready for the classroom the teacher was during the first 9 weeks on the job. The mean ratings for overall readiness and content knowledge were used in the analyses. Beginning teachers were also classified as "ready" or "not ready" and "knowledgeable" or "not knowledgeable" by dichotomizing the rating scale. The analyses comprised one-sample and two-sample t-tests, binomial approximation to the normal distribution, and chi-squared tests of independence. Results provide favorable evidence of criterion-related validity for the two tests, but show no differences between the tests. Results also show that recommended cut-scores would have increased the number of false rejections rather than reducing the number of false acceptances. The results raise questions that require more in-depth examination. An appendix contains study data tables. (Contains 8 tables and 17 references.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

Running head: Evaluating Cut-Scores on Two Certification Tests

ED 431 000

Evaluating Cut-Scores on Two Certification Tests: How Well Do Decisions Based on Cut-Scores Match Teacher- and Principal-Reported Ratings of Competence in the Classroom?

Winifred C. Nweke and Thomas R. Hall

Georgia Professional Standards Commission, Atlanta, Georgia

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
Winifred Nweke
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

TM029826

Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada, April 1999.

BEST COPY AVAILABLE

Abstract

The purpose of this study was three-fold. The first purpose was the investigation of the criterion-related validity evidence for the Georgia Teacher Certification Test (TCT) and Praxis II tests that are used in the teacher certification process in Georgia. The second purpose was to compare decisions based on the two tests. Finally, the effects of using recommended, rather than the adopted, cut-scores were examined. Participants were 2326 beginning teachers in the state of Georgia in the 1998 Fiscal Year as well as their principals. Beginning teachers and their principals completed questionnaires that elicited, on a four-point scale, how well prepared and ready for the classroom the teacher was during the first nine weeks on the job. The mean ratings for overall readiness and content knowledge were used in the analyses. Beginning teachers were also classified as ready or not ready and knowledgeable or not knowledgeable by dichotomizing the rating scale. The analyses comprised one-sample and two-sample t-tests, binomial approximation to the normal distribution as well as chi-squared tests of independence. Results provided favorable evidence of criterion-related validity for the two tests but showed no differences between the tests. Lastly, results showed that recommended cut-scores would have increased the number of false rejections rather than reduce the number of false acceptances. The results raise questions that require more in-depth examination.

Evaluating Cut-Scores on Two Certification Tests: How Well Do Decisions Based on Cut-Scores Match Teacher- and Principal-Reported Ratings of Competence in the Classroom?

Purpose of Study

This study was designed to investigate, first, how well cut-scores on certification tests help in the identification of prospective teachers who rate themselves, and are rated by their principals, knowledgeable in the content area they teach. Specifically, the study provides information to judge the criterion-related validity of the tests. Secondly, this study compares two certification tests, Praxis II and Georgia Teacher Certification Tests (TCT) regarding their ability to identify prospective teachers whose preparation make them feel confident and ready to teach the content assigned. TCT, developed by Georgia Assessment Project and administered by the National Evaluation Systems (NES), was used for certification in the state of Georgia from 1978 to 1997. It was replaced by Praxis II developed and administered by Educational Testing Service (ETS).

Perspective/Theoretical Framework

The certification process is designed to enable the designer to select only individuals who possess enough of the skills or knowledge required to move to the next level or, in the case of employment, individuals who are sufficiently qualified to perform the tasks for which they are being certified. Certification entails verification of some acceptable level of competence. One key ingredient in most certification processes is a test on which a candidate has to perform at or above a cut-score to demonstrate competence. In Georgia, the TCT, Praxis I and II help to assure "minimum basic skills

and subject matter" knowledge for Georgia educators" (Torrey, 1997). The desired level of competence is usually determined in standard-setting exercises using individuals who are deemed knowledgeable in the area of certification. There are numerous methods of setting standards or cut-scores from which to choose, (Angoff, 1971; Ebel, 1972; Jaeger, 1982; Livingston & Zieky, 1982). These are summarized by Jaeger in Linn (1989). More recent modifications of earlier methods include the Bookmark Standard Setting Method proposed by Lewis, Mitzel & Green (1996). But, in the final analysis, as Jaeger concluded, "all standard-setting is judgmental."

Quite often a content validation study precedes standard setting exercise. Evidence of content validity would be sufficient if the test merely measures achievement. Measurement experts disagree on what type of evidence is sufficient or appropriate for a certification test. Content validation of certification or licensing tests based on professional judgment have been upheld in courts in recent times. Consequently, many professional groups or agencies that take this route to validate and set standards of performance on certification tests do not bother to provide evidence of criterion-related validity on the test. Acceptance of content validity has not assuaged the controversy over what type of validity evidence is appropriate on licensing and certification tests. Mehrens (1990) in Mitchell et al (1990) makes a case for content validity while Maddaus (1990) and Mitchell (1990) argue for evidence from all three traditional validity categories: content, criterion-related, and construct validity. Proponents of content validity evidence argue that certification tests do not predict who would be an effective teacher, but rather they eliminate individuals who are not knowledgeable or educated. Opponents, on the other hand, contend that even such a use of certification tests implies that teachers who

have more knowledge make better teachers. Mitchell (1990) is willing to "shift criterion" from trying to select an effective teacher to merely selecting an educated teacher, since that is all that content validation would enable one to say about teacher tests. Mehrens and Lehmann (1991), however, recommend that if a test is used for selection purposes, evidence of its criterion-related validity should be provided, as data become available over and above the content validity evidence.

Ideally, criterion-related validity studies should be conducted before a test is used for selection or certification process in order to avoid the problem of restriction of range. Where the test is in use already, part of the population has been eliminated including individuals who might have performed well on the criterion but had been rejected because they failed on the predictor test. These individuals are called the false negatives or false rejections. The cut-score determines the number of false rejections and false acceptances. The latter are individuals who though they succeeded on the predictor test have not or are not performing adequately on the criterion test or measure. According to Mehrens (1990), many states lower the recommended cut-score to fit with the prevailing political climate, and/or demand or out of concern for lawsuits from false rejections. This process increases the number of false acceptances – individuals who may not be very knowledgeable in the area in which they seek certification. This defeats the purpose of certification tests which are designed to protect the society from incompetent individuals. For the purposes of this study, nothing can be done about the false negatives or rejections who are already eliminated and thus not available for examination. But then, they pose no danger or threat to Georgia's students. False positives, on the other hand, can pose a big danger to the schools. The distribution of false positives or false acceptances will

therefore, be examined. These false acceptances form the second focus of this study. Given the concern over teacher quality (Feldman, 1998) and poor student performance, it is important to assess how well cut-scores on the Praxis II protects the Georgia children from teachers who rate themselves or are rated not ready or competent for Georgia schools.

Method

Data were obtained on 2326 beginning teachers who participated in an earlier study. Some of these had taken the TCT (2239) while others took the Praxis II tests (87) for certification. The TCT was designed to "measure only that content knowledge that teachers themselves judge as essential aspects of classroom teaching" (Georgia Department of Education, 1985). The TCT is comprised of 30 tests. Praxis II: Subject Assessments measure "your knowledge of the subjects you will teach. They also measure your general and subject-specific pedagogical skills and knowledge." (ETS, 1997). Most Praxis II tests are national and a few were written for Georgia. Fifty-three Praxis II tests are used for certification in Georgia.

For psychometric and legal reasons, both TCT and Praxis II tests were validated for use in Georgia before they were adopted. On every test administered in this state, a standard-setting panel recommended a pass score. The test vendors worked with panels, selected by the state agency that is responsible for certification, to determine and recommend cut-scores. Recommended cut-scores are often influenced by the impact such a number would have on pass rates and, in the case of Praxis II, how they compare with pass scores in other states. The cut-scores were either accepted, as recommended, by the Department of Education (for TCT) or by the Commission in charge of professional

standards for teachers, or they were modified before adoption. For example, with regard to TCT and according to a DOE document, "To help ensure equity to examinees, the passing score was set 2.5 standard errors of measurement units (10 percentage points) below the Panels' recommendations on each test." (Department of Education, 1985, p. 6). On some Praxis II tests, the recommended cut-scores are being phased-in in one or two steps over a period of five years (Professional Standards Commission, 1997). This means that cut-scores are set initially at one or two standard errors of measurement below the recommended ones and gradually raised in five years to the recommended scores. This was often used for constructed response tests because the test format is new in the state certification process. Another reason the recommended score might be modified is a result of a cost-benefit analysis of reducing the number of false negatives as opposed to increasing the number of false acceptances. Choosing a lower cut-score than was recommended, together with allowing unlimited retake opportunities for examinees, "virtually eliminates the chance of misclassification of examinees. In short competent examinees have virtually no chance of being classified as not passing" (DOE, 1985, p.6). Phasing in the scores, for whatever reason, allowed the candidates and teacher preparation programs time to adjust to the new test format. The adopted pass score was used to determine whether or not a student passed and hence was eligible for certification.

Beginning teachers' test scores on the certification tests were obtained from the PSC files. To make the numbers comparable, it was planned to convert beginning teachers' scores on these tests into standard error units from the recommended score.

This was not possible due to unavailability of standard errors of measurement for the TCT tests.

After they had been teaching for 9 weeks, beginning teachers completed a survey that elicited, on a 4-point scale, how well prepared and ready for the classroom the teachers considered themselves. The principals or the principals' designates, (from here on also called "principal") similarly completed questionnaires designed to elicit how ready and well prepared they perceived their beginning teachers to be. One of the items required the teacher and the principal to rate how knowledgeable the beginning teacher was in the content area he or she was assigned to teach. Based on the knowledge of content ratings, teachers were classified as knowledgeable (a rating of 3 or 4) or not knowledgeable (a rating of 1 or 2). Similarly, they were classified as ready or not ready overall based on the item that rated overall readiness for the classroom. Beginning teachers who had taken Praxis II tests were also classified as "pass" or "fail" based on the recommended cut-scores. Principals' ratings served as the criterion measure. Teachers' self-ratings on content knowledge and overall readiness were also examined, on an exploratory basis.

Result

Tables 1 and 2 show only half of each corresponding decision table given that the rejected candidates (true and false rejections) are not available for processing. Table 1 shows, based on principals' ratings of teachers' knowledge of content taught, that 4 out of 86 or 4.7% of those who took Praxis II could be classified as false acceptances.

Insert Table 1 about here

Similarly, Table 2 shows 68 participants or 3.2% of those who took TCT could be classified as false acceptances. The corresponding numbers for false acceptances based on teachers' self-ratings are 8.9% and 7.9%, respectively. Success ratios for Praxis II candidates is 95.3% and 96.8% for the TCT based on Principals' rating of content knowledge. Success ratio is determined by dividing the number of beginning teachers that were rated as knowledgeable and ready for the classroom by the total number of candidates selected on each certification test.

Insert Table 2 about here

What is the Impact of Phasing-In Recommended Cut-Scores on False acceptances? This was investigated only for Praxis II since TCT is a much older test and even though there was some documentation of lowering the recommended cut-scores; there was no documentation of phasing them in. When the 87 Praxis II candidates test scores were judged against the recommended score in each test, only six candidates (7%) could be considered false acceptances. Thus, the certification eligibility decision would have remained the same for all the Praxis II candidates except for six teachers. This proportion was significantly different from chance decision ($\underline{P}$ (6 | 87, $\underline{p}$=0.5, $\underline{q}$=0.5) < 0.05).

Also, success ratios based on the recommended cut-scores were determined and compared to those based on the adopted cut-scores. The results are presented in Tables 3 and 4.

Insert Tables 3 and 4 about here

Using the recommended scores, only six beginning teachers would have failed their Praxis II. That would have reduced the number accepted for certification among the beginning teachers to 81 from 87. Only one of the six rated herself or himself as not ready for the classroom. As shown in Tables 3 and 4, there is very little difference between success rates. The biggest difference between the success rates based on the currently adopted and originally recommended scores is 1%. This could be interpreted in two very different ways: The result may be seen as evidence for the justification of lowering the recommended cut-scores in order to keep the false rejections (or false negatives) at a minimum. Another interpretation may wonder whether ensuring that the six candidates are selected is worth creating the impression of lowered certification standards. Proponents of lowered cut-scores will be glad to see that lowering the cut-score does not imply opening the floodgate. They would feel vindicated that five of the six teachers allowed into the profession by this policy feel as ready and knowledgeable as any other teacher.

How knowledgeable and ready for the classroom are teachers who were selected and certified based on their performance on either TCT or Praxis II? This question was examined in two ways. First, one-sample t-test was used to test how ready or knowledgeable the beginning teachers were as rated by teachers themselves and by principals. Thus, four tests of significance were performed with teachers' and principals' ratings on teacher readiness and content knowledge as the four dependent variables. A rating of 3 and above on a 4-point scale were considered ready or knowledgeable. Thus,

the mean readiness and knowledge ratings were compared to 2.99, the cut-off point for non-readiness and inadequacy of knowledge.

The results of the significance tests are presented in Table 5. The results show that the selected teachers were significantly ready and knowledgeable as rated either by themselves or by their principals. Thus, the TCT and Praxis II help select teachers who rate themselves and are rated ready and knowledgeable.

Insert Table 5 about here

The second technique for verifying whether or not the TCT and Praxis II help select ready and knowledgeable teachers was the test of difference of proportions. This was done to test the proportions of teachers in the group who were rated ready or knowledgeable against chance levels ($p = 0.5$). This was important in that the decisions on teacher readiness and knowledgeability are dichotomous rather than continuous. In other words, the critical decision in certification issues was whether or not a given teacher is ready or knowledgeable enough not to pose a threat to students, rather than the group average readiness. Thus, teachers who were rated 1 or 2 on the 4-point scale were considered not ready or not knowledgeable. Similarly, those who were rated 3 or 4 were classified as ready or knowledgeable. As Table 6 shows, teachers rated 95% of themselves as ready while principals rated 92% of the teachers ready.

Insert Table 6 about here

With regard to knowledge of content that the teachers were teaching, 93% of the teachers rated themselves as knowledgeable while principals considered 97% of them as knowledgeable. Using the binomial approximation to the normal distribution, it was

determined that the probability of obtaining the observed proportions by chance was zero. See Table 7.

Insert Table 7 about here

Comparisons of TCT and Praxis II. The second major objective of this study was to compare the two certification tests (TCT and Praxis II) used in Georgia with regard to their ability to select teachers who feel ready and knowledgeable for the classroom and are rated the same by their principals. This was done using t-test for independent samples and chi-squared test of independence. Table 8 shows the mean readiness and content knowledge ratings as assigned by teachers and principals. The only significant effect was the difference between teacher self-ratings on overall readiness of TCT candidates and Praxis II candidates, t (2207) = 2.29, df =2207, p =0.028). Specifically, the teachers did not differ in content knowledge, rated by teachers or by principals, nor did they differ in readiness as rated by principals. The question then is, Is the change from TCT to Praxis justified, especially since the TCT group rated higher on the average on readiness (M = 3.28) than the Praxis II candidates (M= 3.14)?

Insert Table 8 about here

Finally, chi-squared test of independence was used to test the relationship between type of certification test one took and one's classification on readiness and content knowledge. There were no significant relationships between the certification test

taken by a candidate and whether or not the candidate or the principal rated the candidate ready or knowledgeable.

Conclusion

The results of the study provide information that suggests that certification tests used in Georgia show criterion-related validity. They also show that not starting with lower cut-scores than were recommended would have increased the number of false rejections. The results, however, raise the following question: What are the advantages of phasing in the recommended score and appear to begin with lowered standards when, in fact, most candidates would meet the recommended cut-score?

Praxis II is claimed to have better content validity with regard to the content that the grade school teacher in Georgia should know to be able to teach. This study did not find any differences in classroom readiness or reported content knowledge among teachers selected based on the two tests. There were, however, some significant differences in overall readiness, as reported by teachers, in favor of TCT. Further examination of the pattern of responses on the other 24 multiple–choice items and the open-ended questions on the questionnaire may be necessary to see if the two certification tests really select teachers with different characteristics and/or competencies.

References

Angoff, W. H. (1971). Scales, norms, and equivalent scores. In R. L. Thorndike (Ed.), Educational Measurement (pp. 508-600). Washington, DC: American Council on Education.

Ebel, R. L. (1972). Essentials of educational measurement. Englewood Cliffs, NJ: Prentice-Hall.

Educational Testing Service. (1997). The Praxis series: Professional assessments for beginning teachers. Registration Bulletin. Princeton: Author.

Feldman, S. (1998). American Federation of Teachers: Teacher-quality plan; policy on reading teacher preparation and certification. Teacher Education Reports, 20 (15), 4-7.

Georgia Department of Education. (1985). A Standard of quality: The Georgia teacher certification testing program. Atlanta, GA: Author.

Georgia Professional Standards Commission. (1997). Minutes of April 3, 1997 Commission meeting. Atlanta: Author.

Jaeger, R. M. (1982). An iterative structured judgment process for establishing standards on competency tests: Theory and applications. Educational Evaluation and Policy Analysis, 4, 461-475.

Jaeger, R. M. (1989). Certification of student competence, In R. L. Linn. (Ed.), Educational Measurement (3rd ed., pp. 485-514). New York: American Council on Education/Macmillan.

Lewis, D. M., Mitzel, H. C., & Green, D. R. (June 1996). Standard setting: A bookmark approach. In D. R. Green, (Chair), IRT-based standard setting procedures

utilizing behavioral anchoring. Symposium presented at the Council of Chief State School Officers' National Conference on Large-Scale Assessment, Phoenix, AZ.

Livingston, S. A. & Zieky, M. J. (1982). Passing scores: A manual for setting standards on educational and occupational tests. Princeton, NJ: Educational Testing Service.

Maddaus, G. (1990). Legal and professional issues in teacher certification tests. In J. A. Mitchell, S. L. Wise, & B. S. Plake (Eds.), Assessment of teaching: purposes, practices, and implications for the profession. (pp. 209-259). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Mehrens, W. A. (1990). Assessing the quality of teacher assessment tests. In J. A. Mitchell, S. L. Wise, & B. S. Plake (Eds.), Assessment of teaching: purposes, practices, and implications for the profession. (pp. 77-136). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Mehrens, W. A., & Lehmann, I. J. (1991). Measurement and evaluation in Education and Psychology. Chicago: Holt, Rinehart, & Winston.

Mitchell, J. V. (1990). The assessment of teacher assessment: Concluding thoughts and lingering questions. In J. V. Mitchell, S. L. Wise, & B. S. Plake (Eds.), Assessment of teaching: purposes, practices, and implications for the profession. (pp. 347-376). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Mitchell, J. V., Wise, S. L., & Plake, B. S. (Eds.). (1990). Assessment of teaching: purposes, practices, and implications for the profession. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Teacher Education Reports (1998). AFT launches 'Teacher Quality' initiative; Approve policy on reading teacher prep, certification. Washington, D.C.: Author.

Torrey, P. (1997). Report to the Commission on Certification Tests. Atlanta: Professional Standards Commission.

APPENDIX

Table 1. Content Ratings for Praxis II Candidates

	Principals' Ratings	Teachers' Ratings
False Rejections (Not available)	True Acceptances 82	79
True Rejections (Not Available)	False Acceptances 4 (4.7%)	7 (8.1%)
	Total 86	86

Table 2. Content Ratings for TCT Candidates

	Principals' Ratings	Teachers' Ratings
False Rejections (Not available)	True Acceptances 2,271	2,059
True Rejections (Not Available)	False Acceptances 68 (3.2%)	163 (7.3%)
	Total 2,239	2,222

Table 3. Impact of Recommended and Adopted Cut-Scores on Success Ratios for Praxis II Candidates with Rating on Content Knowledge as the Criterion Measure

	PRINCIPALS' RATING			TEACHERS' RATINGS		
	Low	High	Success Rate	Low	High	Success Rate
Based on Adopted Cut-Score	4	82*	95.3%	7	79*	91.9
Based on Recommended Cut-Score	4	76*	95.0%	7	73*	91.3%

* The difference reflects the number of teachers who would have failed if recommended scores were used as cut-scores

Table 4. Impact of Recommended and Adopted Cut-Scores on Success Ratios for Praxis II Candidates with Rating on Overall Readiness as the Criterion Measure

	PRINCIPALS' RATING			TEACHERS' RATINGS		
	Low	High	Success Rate	Low	High	Success Rate
Based on Adopted Cut-Score	11	73*	86.9%	9	76	89.4%
Based on Recommended Cut-Score	11	67*	85.9%	8**	71**	89.9%

*The difference (6) is the number of teachers that would have failed if the recommended cut-score were used.

** One teacher from the low group and five from the high group would have failed if the recommended cut-score were used.

Table 5. One-Sample t-test of Teachers' Mean Readiness and Content Knowledge Ratings

RATER	Readiness	Content Knowledge
TEACHERS	M = 3.27 μ = 2.99 SD = 0.57 df = 2,292 t = 23.82*	M = 3.37 μ = 2.99 SD = 0.65 df = 2,307 t = 28.44*
PRINCIPALS	M = 3.25 μ = 2.99 SD = 0.62 df = 2,218 t = 19.87*	M = 3.37 μ = 2.99 SD = 0.56 df = 2,231 t = 32.30*

- * Significant at $p < 0.05$.

Table 6. Number and Proportions of Beginning Teachers Rated Ready and Knowledgeable

	READINESS		CONTENT KNOWLEDGE	
RATER	Not Ready	Ready	Low Knowledge	High Knowledge
Teachers	121 (5%)	2,172 (95%)	170 (7%)	2,138 (93%)
Principals	177 (8%)	2,042 (92%)	72 (3%)	2,160 (97%)

Table 7. One-Sample Test of Proportions Using the Binomial test

RATER	READINESS	CONTENT KNOWLEDGE
Teachers	x = 2,172 n = 2,293 p = 0.50 P = 0.95 p(x | n, p, q) < 0.01	x = 2,138 n = 2,308 p = 0.50 P = 0.93 p(x | n, p, q) < 0.01
Principals	x = 2,042 n = 2,219 p = 0.50 P = 0.92 p (x | n, p, q) < 0.01	x = 2,160 n = 2,232 p = 0.50 P= 0.97 p(x | n, p, q) < 0.01

Table 8. Comparison between TCT and PRAXIS II Candidates Using Two-Sample Independent t-Test

	READINESS		CONTENT KNOWLEDGE	
RATER	TCT	PRAXIS II	TCT	PRAXIS II
TEACHERS	M = 3.28 SD = 0.57 N = 2,208	M = 3.14 SD = 0.66 N = 85	M = 3.38 SD = 0.65 N = 2,195	M = 3.31 SD = 0.62 N = 85
	t=2.19*, p < 0.05		t = 1.45, p > .05 ns.	
PRINCIPALS	M = 3.25 SD = 0.61 N = 2,135	M = 3.15 SD = 0.70 N = 84	M = 3.37 SD = 0.56 N = 2,117	M = 3.33 SD = 0.61 N = 84
	t= 0.983, p > 0.05, ns.		t= 0.617, p > 0.05, ns.	

* Significant at 0.05 level